DOCUMENT RESUME

ED 450 147 TM 032 337

AUTHOR	Kimbell, Anne-Marie
TITLE	The Basic Concepts of the General Linear Model (GLM): Canonical Correlation Analysis (CCA) as a GLM.
PUB DATE	2001-02-00
NOTE	27p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Correlation; *Multivariate Analysis
IDENTIFIERS	*General Linear Model

ABSTRACT

This paper illustrates how canonical correlation analysis can be used to implement all the parametric tests that canonical methods subsume as special cases. The point is heuristic: all analyses are correlational, apply weights to measured variables to create synthetic variables, and require the interpretation of both weights and structure coefficients. Because all analyses are correlational, "r" squared effect sizes can (and should) be reported in all analyses. An appendix contains the command syntax to run illustrative analyses using the Statistical Package for the Social Sciences (SPSS - v9). (Contains 12 tables and 15 references.) (Author/SLD)

The Basic Concepts of the General Linear Model (GLM):

Canonical Correlation Analysis (CCA) as a GLM

Anne-Marie Kimbell

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 1-3, 2001.

Abstract

The paper illustrates how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases. The point is heuristic: all analyses are correlational, apply weights to measured variables to create synthetic variables, and require the interpretation of both weights and structure coefficients. Because all analyses are correlational, r square effect sizes can (and should) be reported in all analyses.

The Basic Concepts of the General Linear Model (GLM):

Canonical Correlation Analysis (CCA) as a GLM

Jacob "Jack" Cohen (1968) was one of the first to write about the use of regression as a general linear model, establishing that multiple regression subsumes all the univariate parametric analyses of variance techniques. Univariate methods can be used to test hypotheses about the effects of several independent (predictor) variables on a single (dependent) variable, but multivariate methods examine a set of independent variables *and* a set of two or more dependent variables. Several noted researchers have pointed out that this is necessary when conducting research in the behavioral sciences, as multivariate methods both control experimentwise Type I error rate and best honor the reality of the data (Campbell & Taylor, 1996; Fish, 1988; Thompson, 1991, 2000). Experimentwise Type I error rate is limited to the alpha level with multivariate methods because you simultaneously test relationships among all the variables. The reality of the data is best honored with multivariate methods because human behavior involves multiple causes and multiple effects and interactions between multiple variables being studied (Campbell & Taylor, 1996; Campo 1990; Thompson, 2000; Vidal, 1997). Thus Cooley and Lohnes (1971) said canonical correlation analysis "is the simplest model that can begin to do justice to this difficult problem of scientific generalization" (p. 176).

It is also becoming widely understood that canonical correlation analysis is the most general case of the parametric general linear model, subsuming all other parametric univariate and multivariate analyses (Thompson, 1991, 2000). Knapp (1978) wrote that "virtually all the commonly encountered tests of significance can be treated as special cases of canonical correlation analysis" (p. 410). This includes *t* tests, Pearson correlation, analysis of variance [ANOVA], regression, MANOVA, and descriptive discriminant analysis (Campbell & Taylor, 1996; Thompson, 2000). Cohen (1968) noted that while two statistical analyses could yield the same results, a given implementation might provide more useful information or be easier to do. Because the general linear model subsumes all other analyses, it should be used with this in mind. The present paper will illustrate how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases. The point is not that all research ought to be conducted with canonical analyses, but rather the point is a heuristic one: all analyses are correlational; all analyses apply weights to measured variables to create synthetic variables that become the analytic focus; all analyses require the interpretation of both weights and structure coefficients. Furthermore, r square or other effect sizes ought to be reported in every study (Wilkinson & APA Task Force on Statistical Inference, 1999).

The General Linear Model

The general linear model produces an equation that maximizes the relationship of the independent variables to dependent variables. Researchers should understand three important points about the general linear model. The first is that, though the design may be experimental, all analyses are correlational. Experimental design is separate from statistical analysis. Analysis of variance methods are used with the idea that causal inferences may thus be made, but these methods require categorizing variables that should not be categorized, leading to the loss of important data about variance, and do not provide experimental control over these categorized variables unless the design is experimental (Thompson, 1991, 2000).

The second point is that all parametric analyses invoke, either explicitly or implicitly, systems of weights applied to measured variables to create synthetic variables, which are then the focus of the analysis. Thompson (2000) notes that these weights "are often arbitrarily (and confusingly) given different names across different analyses (e.g., beta weights vs. pattern coefficients vs. function coefficients and equation vs. factor vs. function)" (p. 299). These weights, however, are evaluated to determine *what the findings are* rather than if the findings are statistically significant.

The third point is that because all analyses are correlational, they all yield a measure of effect size that is analogous to r^2, which needs to be reported and interpreted. Thompson (2000) suggested that "no knowledgeable researcher reporting bivariate or multiple correlation coefficients fails to comment on the magnitude of the squared correlation coefficient" (p. 299). The 1999 report from the American Psychological Association Task Force on Statistical Inference emphasized that some effect-size estimate should always be provided in every analysis (Wilkinson & APA Task Force on Statistical Inference).

It is important to understand the general linear model principles in order to comprehend that all parametric analyses are related, "facilitating thoughtful researcher judgment in selecting analyses as opposed to employing 'lock-step' decision strategies that limit the utility of analyses" (Henson, 1999, p. 6).

Canonical Correlation Analysis

Canonical correlation analysis is employed to study relationships between two or more variable sets when each set consists of at least two variables. Each set of variables (predictor and criterion) represents a latent construct that the researcher is examining.

Most people use CCA in situations involving only two variable sets, though the analysis can consider more than two sets at a time (Thompson, 2000). The variables must exist within meaningful sets, however, or the use of CCA is not appropriate. The study should involve at least 20 participants per measured variable (Stevens, 1986). If necessary, you can do principal components analysis to compute factor scores to reduce the number of variables.

Because of the complexity of canonical correlation analysis, Thompson (1984) organized some of the research questions that CCA can be used to investigate:

1. To what extent can one set of two or more variables be predicted or "explained" by another set of two or more variables?
2. What contributions does a single variable make to the explanatory power of the set of variables to which the variable belongs?
3. To what extent does a single variable contribute to predicting or "explaining" the composite of the variables in the variable set to which the variable does *not* belong?
4. What different dynamics are involved in the ability of one variable set to "explain" in different ways different portions of the other variable set?
5. What relative power do different canonical functions have to predict or explain relationships?
6. How stable are canonical results across samples or sample subgroups?
7. How closely do obtained canonical results conform to expected canonical results? (p. 10)

Canonical Correlation Analysis as a General Linear Model

In the present analysis, a heuristic data set for 20 elderly persons residing at home, in assisted living, and in nursing homes will be used to demonstrate that canonical correlation subsumes other parametric analyses as special cases. Canonical correlation analysis will be used to perform a t-test, Pearson correlation, multiple regression, ANOVA, MANCOVA, and descriptive discriminant analysis. Table 1 presents heuristic data on four intervally scaled variables related to depression and abuse in the elderly: previous intakes of abuse reports (PREVINT), age (AGE), scores on the Beck Inventory (BECK), and scores on the Indicators of Abuse Screen (IOAS). Also included are grouping data indicating residential location (RESIDE) and gender (GENDER). Five contrast variables are also listed which will be described later.

INSERT TABLE 1 ABOUT HERE

All analyses were run using Statistics for the Social Sciences (SPSS – v9) package. The command syntax for these analyses is included in Appendix A.

The canonical correlation coefficient (R_c) is the correlation between the two sets of synthetic variable scores computed by applying weights to the measured variables. One canonical correlation will be computed for each set of standardized canonical function coefficients and respective measured variables.

Conducting t-test with Canonical Correlation

T-tests are used to determine if the means of two groups are statistically different. A t-test was conducted to determine if the means of males and females (GENDER) differed on the PREVINT variable. Results reported in Table 2 indicate that the difference of the means of the two groups was not statistically significantly different, t = -.138, p = .892. A canonical analysis on the same variables yielded F (1, 18) = .02, p = .892. Table 2 also reports the CCA results, including the canonical correlation (R_c), squared canonical correlation (R_c^2), and Wilks lambda (λ). Wilks lambda, like (R_c^2), is a variance-accounted-for type statistic, but in canonical correlation analysis, it indicates the variance not accounted for (i.e., 1-R_c^2). This lambda is used to test the statistical significance of the canonical correlation (R_c), decreasing (between 0 and 1) as the effect size (R_c^2) increases.

INSERT TABLE 2 ABOUT HERE

The p calculated values are the same in each analysis. The test statistics (t and F) differ only in metric. The F distribution is made up of squared values of the t distribution. Squaring t = -.138 produces .019, which matches the F value of .02. The observed difference in the values is due solely to rounding error by SPSS.

Conducting Pearson Correlation with Canonical Correlation

Pearson correlation (r) is the most frequently used statistic when exploring relationships between two variables. A perfect relationship provides an r = 1 or an r = -1, a perfectly uncorrelated relationship provides an r = 0. The canonical correlation provides the same results, except the canonical is measuring the relationship within multivariate sets.

A Pearson r was computed for PREVINT and AGE. Table 3 reports the obtained results, r = .614, p .004. The canonical correlation analysis computed a squared canonical correlation coefficient of .377. By transforming $\underline{R}_c^2$ =. 377 into $\underline{R}_c$ = .614, the result is identical to the Pearson $\underline{r}$. The $\underline{p}$ values here are also identical. Henson (1999) noted that "Herein lies the most fundamental of general linear model principles: all analyses are correlational. The canonical correlation is nothing more than a bivariate $\underline{r}$ between the synthetic variables created in CCA after the application of weights" (p. 12).

INSERT TABLE 3 ABOUT HERE

Conducting Multiple Regression with Canonical Correlation

Multiple regression uses several variables to predict scores on a criterion variable. In this example, PREVINT was predicted by BECK and IOAS. The SPSS results of the multiple regression and the canonical analysis are presented in Table 4.

INSERT TABLE 4 ABOUT HERE

The squared multiple correlation coefficient ($\underline{R}^2$) derived from the regression analysis was .247, $\underline{F}$ (1, 18) = 2.792, $\underline{p}$ = .089. The canonical analysis resulted in a squared canonical correlation coefficient ($\underline{R}_c^2$) of .247, $\underline{F}$ (1, 18) = 2.7916, $\underline{p}$ = .089. Rounding by the computer package accounts for any difference in values. Note that Beta weights ($\underline{B}$) and standardized function coefficients are easily converted into each other using the following formulas:

$\underline{B}$ / $\underline{R}_c$ = Function Coefficient

Function Coefficient * $\underline{R}$ = $\underline{B}$

For example, BECK had a $\underline{B}$ weight of -.048. Using $\underline{R}_c$ = .497 from the CCA, we find that the standardized function coefficient matches, within rounding error, that reported in

Table 4 (-.048 / .497 = -.096). With these formulas and because we know that the regression multiple R equals the canonical R_c, we can find canonical function coefficients using only a regression analysis and find B weights using only canonical correlation analysis.

Conducting Factorial ANOVA with Canonical Correlation

Table 1 included five orthogonal contrast variables that were created with SPSS commands (see syntax file in Appendix A). Analysis of variance methods use planned contrasts to test specific, theory-driven hypotheses against omnibus hypotheses (Thompson, 1994). They are presented here to show that canonical correlation analysis can conduct ANOVA.

A 3 X 2 factorial ANOVA was conducted with GENDER and RESIDE as independent variables and PREVINT as the dependent variable. For the CCA, the contrast variables from Table 1 were used. The total number of contrasts needed to carry out an ANOVA equals the degrees of freedom for each main effect. The RESIDE main effect has two degrees of freedom and is represented by CRE1 and CRE2. The GENDER main effect is represented by CGENDER with one degree of freedom. CGRRE1 and CGRRE2 are cross products of the other main effects and test the RESIDE X GENDER interaction effects. Table 5 presents results for the ANOVA: RESIDE, $F = 3.168$; GENDER, $F = .051$; RESIDE X GENDER, $F = .563$. The error effect for the full ANOVA model, .664981, was computed by dividing the sum of squares error by the sum of squares total (131.500 / 197.750).

INSERT TABLE 5 ABOUT HERE

The canonical analysis is conducted in a series of steps, beginning with the creation of four separate designs, using PREVINT as the dependent measure and the contrasts as independent variables. Design 1 included all planned contrasts, CRE1, CRE2, CGENDER, CGRRE1, and CGRRE2 to test the total effect (SOS explained). Design 2 used CGENDER, CGRRE1, and CGRRE2 to jointly test the GENDER and interaction effects. Design 3 used CRE1, CRE2, CGRRE1, and CGRRE2 to jointly test the RESIDE and interaction effects. The final CCA, Design 4, used CRE1, CRE2, and CGENDER to jointly test the RESIDE and GENDER effects. Table 6 displays the Wilks lambda values for each design. Thompson (1994) noted that lambda is analogous to a sum of squares in ANOVA and is a "reverse" effect size, equaling the effect for the error term. Comparing the $\underline{\lambda}$ = .66498 for the total effect (Table 6) with the error effect size (sum of squares error / sum of squares total: 131.500 / 197.750 = .664981) (Table 5) confirms this relationship between the statistics.

INSERT TABLE 6 ABOUT HERE

The next step is to convert the canonical lambdas to separate omnibus ANOVA effects by dividing the total effect lambda by the lambda value for each design (effect). To compute the ANOVA lambda for the RESIDE main effect, the total lambda (.66498) was divided by the Design 2 lambda (.96590), which reflects the joint effect of the contrast variables for the GENDER main effect and the RESIDE X GENDER interaction effect. This process "removes" the effect of the other hypotheses, leaving the omnibus lambda for the RESIDE main effect to be .6884564 (.66498 / .96590 = .6884564 = $\underline{\lambda}$). The same process was then used to find the other ANOVA lambdas with results reported in Table 7.

INSERT TABLE 7 ABOUT HERE

The final step is to convert ANOVA lambdas into ANOVA F statistics using the following formula:

[(1 – Lambda) / Lambda] * (df error / df effect) = F

To illustrate, the F value for the RESIDE main effect was modeled by [(1 - .6884564) / .6884564] * (14 / 2) = 3.168. Table 8 includes the transformations for the main effects and the interaction. Notice that the F calculations are the same as the ANOVA F calculations in Table 5.

INSERT TABLE 8 ABOUT HERE

Conducting Factorial MANOVA with Canonical Correlation

A 3 X 2 factorial MANOVA was calculated using PREVINT and AGE as dependent variables and RESIDE and GENDER as independent variables. Results from this analysis are found in Table 9. As with the ANOVA calculations above, four CCA designs using the contrast variables were run with the canonical lambdas reported in Table 10. Table 11 contains the conversion of the canonical lambdas into MANOVA lambdas. Note the equivalence of the MANOVA $\underline{\lambda}$s in Table 9 with those obtained through the canonical analysis in Table 11. The final conversion to F values was not required here as MANOVA uses the λ value to calculate F statistics, as against the SOS value in ANOVA.

INSERT TABLES 9 – 11 ABOUT HERE

Conducting Discriminant Analysis with Canonical Correlation

Discriminant analysis techniques can either be used predictively to classify persons into groups or descriptively where variables identify latent structures among groups (Huberty, 1994). This analysis was conducted with GENDER as the nominally scaled predictor variable and PREVINT and AGE as criterion variables. Table 12 reports a non-statistically significant result X^2 (2,17) of .149, p = .928. The canonical analysis was conducted using the planned contrast variable CGENDER as the predictor. Results of the CCA are also reported in Table 12. Note that the results are identical for the two analyses. The reporting of the X^2 and F statistics are the only difference, but these are arbitrary, as they represent the same value expressed in a different metric.

INSERT TABLE 12 ABOUT HERE

Conclusion

The purpose of the present paper has been to demonstrate that canonical correlation analysis subsumes all other parametric analytic methods and is, therefore, the most general case of the general linear model. Researchers should be selective in the methods they use for analysis, avoiding the mistake of discarding variance in data when using OVA methods with nominally scaled variables, and using CCA when appropriate.

References

Arnold, M. (1996, January). The relationship of canonical correlation analysis to other parametric methods. Paper presented at the annual meeting of the Southwestern Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 395 994)

Campbell, K. T., & Taylor, D. L. (1996). Canonical correlation analysis as a general linear model: A heuristic lesson for teachers and students. Journal of Experimental Education, 64, 157-171.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cooley, W. W. & Lohnes, P. R. (1976). Evaluation research in education. New York: Irvington.

Fish, L. J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Henson, R. K. (1999, January). An illustration that there is a multivariate parametric general linear model: Canonical correlation analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio. (ERIC Document Reproduction Service No. ED 428 081)

Huberty, C. (1994). Applied discriminant analysis. New York: Wiley.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahway, NJ: Erlbaum.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park, CA: Sage.

Thompson, B. (1991) Methods, plainly speaking: A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24, 80-94.

Thompson, B. (1994). Planned versus unplanned and orthogonal versus nonorthogonal contrasts: The neo-classical perspective. In B. Thompson (Ed.), Advances in social science methodology (Vol. 3, pp. 3-27). Greenwich, CT: JAI Press.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 285-316). Washington, DC: American Psychological Association.

Vidal, S. (1997, January). Canonical Correlation Analysis as the General Linear Model. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 407 428)

Wilkinson, L., & APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604. (Available online http://www.apa.org/journals/amp/amp548594.html)

TABLE 1

Heuristic Data (n = 20) for Canonical Correlation Analysis Illustration

ID	PREVINT	AGE	BECK	IOAS	RESIDE	GENDER	CRE1	CRE2	CGENDER	CGRRE1	CGRRE2
1	1	69	55	27	1	2	-1	-1	1	-1	-1
2	2	72	20	17	1	1	-1	-1	-1	1	1
3	8	90	49	5	3	2	1	-1	1	1	-1
4	3	66	38	12	2	2	0	2	1	0	2
5	4	69	17	14	1	1	-1	-1	-1	1	1
6	5	65	26	16	2	2	0	2	1	0	2
7	7	71	45	10	3	2	1	-1	1	1	-1
8	3	78	14	25	2	2	0	2	1	0	2
9	8	90	13	24	2	1	0	2	-1	0	-2
10	12	98	52	5	3	1	1	-1	-1	-1	1
11	4	77	11	22	1	2	-1	-1	1	-1	-1
12	7	73	12	1	2	2	0	2	1	0	2
13	3	66	20	12	1	1	-1	-1	-1	1	1
14	9	62	9	11	1	1	-1	-1	-1	1	1
15	11	81	8	14	2	2	0	2	1	0	2
16	6	83	47	11	3	1	1	-1	-1	-1	1
17	9	85	16	17	2	2	0	2	1	0	2
18	4	67	5	18	1	2	-1	-1	1	-1	-1
19	8	69	44	8	3	2	1	-1	1	1	-1
20	1	62	13	5	2	1	0	2	-1	0	-2

TABLE 2

Conducting t-test with Canonical Correlation (PREVINT BY GENDER).

t-test Analysis		Canonical Analysis	
$t(18)$	-.138	F (1, 18)	.02
p	.892	p	.02
M (GENDER 1)	8		
SD	3.81	R_c	.031
M (GENDER 2)	12	R_c^2	.001
SD	5.83	lambda	.999

TABLE 3

Conducting Pearson Correlation with Canonical Correlation (PREVINT BY AGE)

Pearson r Analysis		Canonical Analysis	
r	.614	R_c	.614
		R_c^2	.377
		lambda	.623
p	.004	P	.004

TABLE 4

Conducting Multiple Regression with Canonical Correlation (PREVINT by BECK and AGE).

Multiple Regression Analysis		Canonical Analysis	
R	.497	R_c	.497
R^2	.247	R_c^2	.247
F (1, 18)	2.792	F (1, 18)	2.7926
p	.089	p	.089
		lambda	.315
Beta Weights		Function Coefficients	
BECK	-.048	.096	
IOAS	-.508	1.022	

TABLE 5

3 X 2 Factorial ANOVA (PREVINT by RESIDE by GENDER).

Source	SOS	df	MS	F	p	r^2
RESIDE	59.506	2	29.753	3.168	.073	30.09%
GENDER	.480	1	.480	.051	.824	.24%
R X G	10.574	2	5.287	.563	.582	5.35%
Error	131.500	14	9.393	1.411		
Total	197.750	9	10.408			

TABLE 6

Canonical Analysis on Four Designs (PREVINT).

Design	Independent Variables	lambda
1	CRE1, CRE2, CGENDER, CGRRE1, CGRRE2	.66498
2	CGENDER, CGRRE1, CGRRE2	.96590
3	CRE1, CRE2, CGRRE1, CGRRE2	.66741
4	CRE1, CRE2, CGENDER	.71845

TABLE 7

Conversion of Canonical Lambdas to Omnibus ANOVA Lambdas.

ANOVA Effect	Designs	Transformation	ANOVA lambda
RESIDE	1 / 2	.66498 / .96590	.688456
GENDER	1 / 3	.66498 / .66741	.99636
RESIDE X GENDER	1 / 4	.66498 / .71845	.92558

TABLE 8

Conversion of ANOVA Lambdas to ANOVA F Statistics

Source	Transformation	F
RESIDE	[(1 - .688456) / .688456] * (14 / 2) =	3.168
GENDER	[(1-.99636) / .99636] * (14 / 1) =	.051
RESIDE X GENDER	[(1-.92558) / .92558] * (14 / 2) =	.563

TABLE 9

3 X 2 Factorial MANOVA (PREVINT and AGE by RESIDE and GENDER).

Source	lambda	df	F	p
RESIDE	.61612	4, 26	1.78098	.163
GENDER	.94694	2, 13	.36421	.702
RESIDE X GENDER	.73343	4, 26	1.08985	.382

TABLE 10

Canonical Analysis on Four Designs (PREVINT and AGE by Contrasts).

Design	Independent Variables	lambda
1	CRE1, CRE2, CGENDER, CGRRE1, CGRRE2	.47617
2	CGENDER, CGRRE1, CGRRE2	.77285
3	CRE1, CRE2, CGRRE1, CGRRE2	.50285
4	CRE1, CRE2, CGENDER	.64923

TABLE 11

Conversion of Canonical Lambdas to Omnibus MANOVA Lambdas.

MANOVA Effect	Designs	Transformation	MANOVA lambda
RESIDE	1 / 2	.47617 / .77285	.6161222
GENDER	1 / 3	.47617 / .50285	.9469424
RESIDE X GENDER	1 / 4	.47617 / .64923	.7334381

TABLE 12

Conducting Discriminant Analysis with Canonical (PREVINT and AGE by GENDER).

Discriminant Analysis		Canonical Analysis	
R_c	.093	R_c	.093
R_c^2	.0086	R_c^2	.009
lambda	.991	lambda	.991
X^2	.149	F	.0746
df	2, 17	df	2, 17
p	.928	p	.928

Appendix A

```
SET BLANKS=SYSMIS UNDEFINED=WARN printback listing.
TITLE 'Canonical Correlation Analysis as the General Linear Model' .
COMMENT****************************************************
COMMENT Heuristic data for 20 cases
COMMENT PREVINT - previous reports of abuse
COMMENT AGE - age
COMMENT BECK - Beck Depression Inventory
COMMENT IOAS - Indicators of Abuse scale
COMMENT RESIDE - home(1), assisted living(2), nursing home(3)
COMMENT GENDER - male(1), female(2) .
DATA LIST
 FILE='a:ccaglm1.txt' FIXED RECORDS=1/
 ID 1-2 PREVINT 4-5 AGE 7-8 BECK 10-11 IOAS 13-14 RESIDE 16
 GENDER 18 .
EXECUTE.
list variables=all/cases=999/format=numbered .

COMMENT show that cca can do t-test.
T-TEST
 GROUPS=GENDER(1 2)
 /MISSING=ANALYSIS
 /VARIABLES=PREVINT
 /CRITERIA=CIN (.95) .
MANOVA
 GENDER WITH PREVINT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM= (STAN ESTIM COR) .
COMMENT Show that cca can do Pearson r.
CORRELATIONS
 /VARIABLES=PREVINT AGE
 /PRINT=TWOTAIL NOSIG
 /MISSING=PAIRWISE .
MANOVA
 PREVINT WITH AGE
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
COMMENT Show that cca can do multiple regression.
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN (.05) POUT (.10)
 /NOORGIN
 /DEPENDENT PREVINT
 /METHOD=ENTER BECK IOAS .
MANOVA
 BECK IOAS WITH PREVINT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
COMMENT Show that cca can do factorial ANOVA
```

```
COMMENT Compute contrast variables to do cca.
IF (RESIDE = 1) CRE1 = -1 .
IF (RESIDE = 2) CRE1 = 0 .
IF (RESIDE = 3) CRE1 = 1 .
COMMENT Tests equality of the means of home(7) vs. nursing home (5) residence.
EXECUTE .
IF (CRE1 = -1) CRE2 = -1 .
IF (CRE1 = 0) CRE2 = 2 .
IF (CRE1 = 1) CRE2 = -1 .
EXECUTE .
COMMENT Tests equality of means of assisted living(8) vs. home and nursing home(12)
residence .
IF (GENDER = 1) CGENDER = -1 .
IF (GENDER = 2) CGENDER = 1 .
EXECUTE .
COMMENT Tests equality of means of males (8) vs. females (12) .
COMPUTE CGRRE1 = CRE1 * CGENDER .
COMPUTE CGRRE2 = CRE2 * CGENDER .
EXECUTE .
COMMENT Tests gender by residence effects.
COMMENT Show contrast variables are orthogonal .
CORRELATIONS
 /VARIABLES=CRE1 CRE2 CGENDER CGRRE1 CGRRE2
 /PRINT=TWOTAIL SIG
 /MISSING=PAIRWISE .
COMMENT Step one: run factorial ANOVA and cca on construct variables.
ANOVA
 VARIABLES=PREVINT
 BY RESIDE(1 3) GENDER(1 2)
 /MAXORDERS ALL
 /METHOD UNIQUE
 /FORMAT LABELS .
MANOVA
 CRE1 CRE2 CGENDER CGRRE1 CGRRE2 WITH PREVINT
 /PRING=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
MANOVA
 CGENDER CGRRE1 CGRRE2 WITH PREVINT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
MANOVA
 CRE1 CRE2 CGRRE1 CGRRE2 WITH PREVINT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
MANOVA
 CRE1 CRE2 CGENDER WITH PREVINT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
COMMENT Show cca can do MANOVA.
MANOVA
 PREVINT AGE BY RESIDE (1 3) GENDER(1 2)
 /PRINT SIGNIF(MULT UNIV)
 /NOPRINT PARAM (ESTIM)
 /METHOD=UNIQUE
 /ERROR WITHIN+RESIDUAL
 /DESIGN .
```

```
MANOVA
 CRE1 CRE2 CGENDER CGRRE1 CGRRE2 WITH PREVINT AGE
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
MANOVA
 CGENDER CGRRE1 CGRRE2 WITH PREVINT AGE
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
MANOVA
 CRE1 CRE2 CGRRE1 CGRRE2 WITH PREVINT AGE
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
MANOVA
 CRE1 CRE2 CGENDER WITH PREVINT AGE
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
COMMENT Show cca can do discriminant analysis.
DISCRIMINANT
 /GROUPS=GENDER (1 2)
 /VARIABLES=PREVINT AGE
 /ANALYSIS ALL
 /PRIORS EQUAL
 /CLASSIFY=NONMISSING POOLED .
MANOVA
 PREVINT AGE WITH CGENDER
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR) .
```

TM032337

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: THE BASIC CONCEPTS OF THE GENERAL LINEAR MODEL (GLM): CANONICAL CORRELATION ANALYSIS (CCA) AS A GLM

Author(s): Anne-Marie Kimbell

Corporate Source:	Publication Date: 2/1/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document		Sample sticker to be affixed to document [ ]	
Check here Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction	"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY ANNE-MARIE KIMBELL TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."	"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."	**or here** Permitting reproduction in other than paper copy.
	Level 1	Level 2	

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X Anne-Marie Kimbell	Position: RES ASSOCIATE
Printed Name: ANNE-MARIE KIMBELL	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: 979 (~~409~~)
	Date: 1/18/01

OVER